

Model Reduction using a Frequency-Limited \mathcal{H}_2 -Cost

Daniel Petersson^{a*} and Johan Löfberg^a

^a Division of Automatic Control,
Department of Electrical Engineering,
Linköpings universitet, SE-581 83 Sweden

December 10, 2012

Abstract

We propose a method for model reduction on a given frequency range, without the use of input and output filter weights. The method uses a nonlinear optimization approach to minimize a frequency limited \mathcal{H}_2 like cost function. An important contribution in the paper is the derivation of the gradient of the proposed cost function. The fact that we have a closed form expression for the gradient and that considerations have been taken to make the gradient computationally efficient to compute enables us to efficiently use off-the-shelf optimization software to solve the optimization problem.

1 Introduction

Given a linear time-invariant (LTI) dynamical model,

$$\begin{aligned}\dot{x}(t) &= \mathbf{A}x(t) + \mathbf{B}u(t), \\ y(t) &= \mathbf{C}x(t) + \mathbf{D}u(t)\end{aligned}$$

where $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\mathbf{B} \in \mathbb{R}^{n \times m}$, $\mathbf{C} \in \mathbb{R}^{p \times n}$ and $\mathbf{D} \in \mathbb{R}^{p \times m}$, the model reduction problem is to find a reduced order model

$$\begin{aligned}\dot{x}_r(t) &= \mathbf{A}_r x_r(t) + \mathbf{B}_r u(t), \\ y_r(t) &= \mathbf{C}_r x_r(t) + \mathbf{D}_r u(t),\end{aligned}$$

with $\mathbf{A}_r \in \mathbb{R}^{n_r \times n_r}$, $\mathbf{B}_r \in \mathbb{R}^{n_r \times m}$, $\mathbf{C}_r \in \mathbb{R}^{p \times n_r}$ and $\mathbf{D}_r \in \mathbb{R}^{p \times m}$ with $n_r < n$, where this reduced order model describes the original model well in some metric. In this paper we are interested in a reduced order model that describes the model well on a given frequency range. This is motivated by situations where the given model is valid only for a certain frequency range, for example as in

*Corresponding author. Email: petersson@isy.liu.se

[9] where models coming from aerodynamical and structural mechanics computations describing a flexible structure are only valid up to a certain frequency.

For a review of model reduction approaches, both ordinary and frequency-weighted, see e.g. [4] and [3]. Some of the most commonly used frequency-weighted methods, according to [4], are [10], [1] and [6], which all use different balanced truncation approaches. In many of the frequency-weighted methods one has to specify input and output filter weights. In [2] they introduce a method which does not need these weighting functions, by introducing frequency-limited Gramians. This method can be interpreted as using ideal low-, band- or high-pass filters as weights. However, this method has the drawback of not always producing stable models. One approach to remedy this has been presented in [4], where they introduce a modification of the method in [2], and additionally derive an \mathcal{H}_∞ bound for the error. [8] also presents a modification of the method from [2], however this method is only applicable to SISO models. Note that these methods, whilst having an upper bound on the \mathcal{H}_∞ error, do not minimize an explicit measure.

One of the main contributions in this paper is a method which uses optimization, and not truncation, to find an \mathcal{H}_2 -optimal reduced order model, and, where this optimization is only performed over a limited frequency interval. Another important contribution is the derivation of the gradient for the cost function. The fact that we have a closed form expression for the gradient and that considerations have been taken to make the gradient computationally efficient to compute enables us to efficiently use off-the-shelf optimization software to solve the optimization problem.

2 Frequency Limited Gramians

The method proposed in this paper uses the idea presented in [2], in which they introduce frequency-limited Gramians. Before introducing the frequency-limited Gramians we define the standard Gramians, see [11], in time and frequency domain, for reference. For a system G , that is stable and described by

$$G : \begin{cases} \dot{x}(t) &= \mathbf{A}x(t) + \mathbf{B}u(t) \\ y(t) &= \mathbf{C}x(t) + \mathbf{D}u(t) \end{cases} \quad (1)$$

denoted as $G : \left[\begin{array}{c|c} \mathbf{A} & \mathbf{B} \\ \hline \mathbf{C} & \mathbf{D} \end{array} \right]$, the observability and controllability Gramians are defined as

$$\mathbf{P} = \int_0^\infty e^{\mathbf{A}\tau} \mathbf{B} \mathbf{B}^\top e^{\mathbf{A}^\top \tau} d\tau = \frac{1}{2\pi} \int_{-\infty}^\infty \mathbf{H}(\nu) \mathbf{B} \mathbf{B}^\top \mathbf{H}^*(\nu) d\nu, \quad (2a)$$

$$\mathbf{Q} = \int_0^\infty e^{\mathbf{A}^\top \tau} \mathbf{C}^\top \mathbf{C} e^{\mathbf{A}\tau} d\tau = \frac{1}{2\pi} \int_{-\infty}^\infty \mathbf{H}^*(\nu) \mathbf{C}^\top \mathbf{C} \mathbf{H}(\nu) d\nu, \quad (2b)$$

where $\mathbf{H}(\omega) = (i\omega\mathbb{I} - \mathbf{A})^{-1} \mathbf{B}$ and $\mathbf{H}^*(\omega)$ denotes the conjugate transpose of $\mathbf{H}(\omega)$. The controllability and observability Gramians satisfy, respectively, the Lyapunov equations

$$\mathbf{A} \mathbf{P} + \mathbf{P} \mathbf{A}^\top + \mathbf{B} \mathbf{B}^\top = \mathbf{0}, \quad (3a)$$

$$\mathbf{A}^\top \mathbf{Q} + \mathbf{Q} \mathbf{A} + \mathbf{C}^\top \mathbf{C} = \mathbf{0}. \quad (3b)$$

Now we narrow the frequency band, from $(-\infty, \infty)$ to $(-\omega, \omega)$ where $\omega < \infty$. We define the frequency-limited Gramians, see [2], as

$$\mathbf{P}_\omega = \frac{1}{2\pi} \int_{-\omega}^{\omega} \mathbf{H}(\nu) \mathbf{B} \mathbf{B}^\top \mathbf{H}^*(\nu) d\nu, \quad (4a)$$

$$\mathbf{Q}_\omega = \frac{1}{2\pi} \int_{-\omega}^{\omega} \mathbf{H}^*(\nu) \mathbf{C}^\top \mathbf{C} \mathbf{H}(\nu) d\nu. \quad (4b)$$

These Gramians can be shown to satisfy the following Lyapunov equations, see [2],

$$\mathbf{A} \mathbf{P}_\omega + \mathbf{P}_\omega \mathbf{A}^\top + \mathbf{S}_\omega \mathbf{B} \mathbf{B}^\top + \mathbf{B} \mathbf{B}^\top \mathbf{S}_\omega^* = \mathbf{0}, \quad (5a)$$

$$\mathbf{A}^\top \mathbf{Q}_\omega + \mathbf{Q}_\omega \mathbf{A} + \mathbf{S}_\omega^* \mathbf{C}^\top \mathbf{C} + \mathbf{C}^\top \mathbf{C} \mathbf{S}_\omega = \mathbf{0}, \quad (5b)$$

with

$$\mathbf{S}_\omega = \frac{i}{2\pi} \ln ((\mathbf{A} + i\omega \mathbb{I})(\mathbf{A} - i\omega \mathbb{I})^{-1}) \quad (6)$$

In [2] they continue by creating a balanced system and performing a balanced truncation using the newly defined frequency-limited Gramians. A drawback with this method is that since the terms $\mathbf{S}_\omega \mathbf{B} \mathbf{B}^\top + \mathbf{B} \mathbf{B}^\top \mathbf{S}_\omega^*$ and $\mathbf{S}_\omega^* \mathbf{C}^\top \mathbf{C} + \mathbf{C}^\top \mathbf{C} \mathbf{S}_\omega$ are not guaranteed to be positive definite, stability of the reduced order model cannot be guaranteed. There exist a modification to this in [4] where they propose a remedy to this.

Remark 1. *By using addition/subtraction of two or more different frequency-limited Gramians it is possible to focus on one or more arbitrary frequency ranges, e.g., you can construct the frequency-limited controllability Gramian, \mathbf{P}_Ω , for the interval $\omega \in \Omega = [\omega_1, \omega_2] \cup [\omega_3, \omega_4]$ as*

$$\mathbf{A} \mathbf{P}_\Omega + \mathbf{P}_\Omega \mathbf{A}^\top + \mathbf{S}_\Omega \mathbf{B} \mathbf{B}^\top + \mathbf{B} \mathbf{B}^\top \mathbf{S}_\Omega^* = \mathbf{0}, \quad (7)$$

with $\mathbf{S}_\Omega = \mathbf{S}_{\omega_2} - \mathbf{S}_{\omega_1} + \mathbf{S}_{\omega_4} - \mathbf{S}_{\omega_3}$.

3 Frequency Limited Model Reduction using Optimization

The \mathcal{H}_2 -norm of G , in (1), can be expressed as

$$\|G\|_{\mathcal{H}_2}^2 = \text{tr} \int_0^\infty \mathbf{C} \mathbf{e}^{\mathbf{A}\tau} \mathbf{B} \mathbf{B}^\top \mathbf{e}^{\mathbf{A}^\top \tau} \mathbf{C}^\top d\tau \quad (8a)$$

$$= \frac{1}{2\pi} \text{tr} \int_{-\infty}^\infty G(i\nu) G^*(i\nu) d\nu \quad (8b)$$

$$= \frac{1}{2\pi} \text{tr} \int_{-\infty}^\infty \mathbf{C} \mathbf{H}(\nu) \mathbf{B} \mathbf{B}^\top \mathbf{H}^*(\nu) \mathbf{C}^\top d\nu = \text{tr} \mathbf{C} \mathbf{P} \mathbf{C}^\top \quad (8c)$$

$$= \text{tr} \int_0^\infty \mathbf{B}^\top \mathbf{e}^{\mathbf{A}^\top \tau} \mathbf{C}^\top \mathbf{C} \mathbf{e}^{\mathbf{A}\tau} \mathbf{B} d\tau \quad (8d)$$

$$= \frac{1}{2\pi} \text{tr} \int_{-\infty}^\infty \mathbf{B}^\top \mathbf{H}^*(\nu) \mathbf{C}^\top \mathbf{C} \mathbf{H}(\nu) \mathbf{B} d\nu = \text{tr} \mathbf{B}^\top \mathbf{Q} \mathbf{B}. \quad (8e)$$

In this paper we introduce a new frequency-limited \mathcal{H}_2 -like norm that uses the frequency-limited Gramians presented in the previous section, and we denote the new measure by $\|G\|_{\mathcal{H}_2, \omega}$, with

$$\|G\|_{\mathcal{H}_2, \omega}^2 = \frac{1}{2\pi} \operatorname{tr} \int_{-\omega}^{\omega} G(i\nu) G^*(i\nu) d\nu \quad (9)$$

$$= \frac{1}{2\pi} \operatorname{tr} \int_{-\omega}^{\omega} (\mathbf{C}\mathbf{H}(\nu)\mathbf{B} + \mathbf{D}) (\mathbf{B}^\top \mathbf{H}^*(\nu) \mathbf{C}^\top + \mathbf{D}^\top) d\nu \quad (10)$$

$$= \operatorname{tr} \mathbf{C} \mathbf{P}_\omega \mathbf{C}^\top + 2 \operatorname{tr} \left[\left(\mathbf{C} \mathbf{S}_\omega \mathbf{B} + \mathbf{D} \frac{\omega}{2\pi} \right) \mathbf{D}^\top \right]. \quad (11)$$

One thing that differs from the ordinary \mathcal{H}_2 -norm is that, if we do not include an infinite interval in Ω , i.e., include $\omega = \infty$ as the end frequency, then the system does not need to be strictly proper. This means that we can, in this case, have $\mathbf{D} \neq \mathbf{0}$.

The method proposed in this paper is a model reduction method that, given a model G , finds a reduced order model, \hat{G} , that is a good approximation on a given frequency interval, e.g. $[0, \omega]$. The objective is to minimize the error between the given model and the sought reduced order model in a frequency-limited \mathcal{H}_2 -norm, using the frequency-limited Gramians. We formulate the optimization problem

$$\underset{\hat{G}}{\text{minimize}} \left\| G - \hat{G} \right\|_{\mathcal{H}_2, \omega}^2 = \underset{\hat{G}}{\text{minimize}} \|E\|_{\mathcal{H}_2, \omega}^2, \quad (12)$$

where

$$\|E\|_{\mathcal{H}_2, \omega}^2 = \frac{1}{2\pi} \operatorname{tr} \int_{-\omega}^{\omega} E(i\nu) E^*(i\nu) d\nu. \quad (13)$$

Assume that the system E is stable and described by

$$E : \left[\begin{array}{c|c} \mathbf{A}_E & \mathbf{B}_E \\ \hline \mathbf{C}_E & \mathbf{D}_E \end{array} \right]. \quad (14)$$

Given G and \hat{G} , represented as

$$G : \left[\begin{array}{c|c} \mathbf{A} & \mathbf{B} \\ \hline \mathbf{C} & \mathbf{D} \end{array} \right], \hat{G} : \left[\begin{array}{c|c} \hat{\mathbf{A}} & \hat{\mathbf{B}} \\ \hline \hat{\mathbf{C}} & \hat{\mathbf{D}} \end{array} \right], \quad (15)$$

the error system can be realized, in state space form, as

$$E : \left[\begin{array}{c|c} \mathbf{A}_E & \mathbf{B}_E \\ \hline \mathbf{C}_E & \mathbf{D}_E \end{array} \right] = \left[\begin{array}{c|c} \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{A}} \end{pmatrix} & \begin{pmatrix} \mathbf{B} \\ \hat{\mathbf{B}} \end{pmatrix} \\ \hline \begin{pmatrix} \mathbf{C} & -\hat{\mathbf{C}} \end{pmatrix} & \mathbf{D} - \hat{\mathbf{D}} \end{array} \right]. \quad (16)$$

This realization of the error system will later prove beneficial when rewriting the optimization problem. Throughout the paper we will assume that the given model is stable.

The cost function for the optimization problem (12) can be written as

$$\|E\|_{\mathcal{H}_2, \omega}^2 = \operatorname{tr} \mathbf{C}_E \mathbf{P}_{E, \omega} \mathbf{C}_E^\top + 2 \operatorname{tr} \left[\left(\mathbf{C}_E \mathbf{S}_{E, \omega} \mathbf{B}_E + \mathbf{D}_E \frac{\omega}{2\pi} \right) \mathbf{D}_E^\top \right] \quad (17a)$$

$$= \operatorname{tr} \mathbf{B}_E^\top \mathbf{Q}_{E, \omega} \mathbf{B}_E + 2 \operatorname{tr} \left[\left(\mathbf{C}_E \mathbf{S}_{E, \omega} \mathbf{B}_E + \mathbf{D}_E \frac{\omega}{2\pi} \right) \mathbf{D}_E^\top \right]. \quad (17b)$$

where

$$\mathbf{A}_E \mathbf{P}_{E,\omega} + \mathbf{P}_{E,\omega} \mathbf{A}_E^\top + \mathbf{S}_{E,\omega} \mathbf{B}_E \mathbf{B}_E^\top + \mathbf{B}_E \mathbf{B}_E^\top \mathbf{S}_{E,\omega}^* = \mathbf{0}, \quad (18a)$$

$$\mathbf{A}_E^\top \mathbf{Q}_{E,\omega} + \mathbf{Q}_{E,\omega} \mathbf{A}_E + \mathbf{S}_{E,\omega}^* \mathbf{C}_E^\top \mathbf{C}_E + \mathbf{C}_E^\top \mathbf{C}_E \mathbf{S}_{E,\omega} = \mathbf{0}, \quad (18b)$$

with

$$\mathbf{S}_{E,\omega} = \frac{i}{2\pi} \ln \left((\mathbf{A}_E + i\omega \mathbb{I})(\mathbf{A}_E - i\omega \mathbb{I})^{-1} \right). \quad (19)$$

In this paper we have also derived a simpler expression for the matrix $\mathbf{S}_{E,\omega}$, compared to what is presented in [2].

Lemma 1. *For a matrix \mathbf{A} that is Hurwitz we have that*

$$\mathbf{S}_\omega = \frac{i}{2\pi} \ln \left((\mathbf{A} + i\omega \mathbb{I})(\mathbf{A} - i\omega \mathbb{I})^{-1} \right) = \text{Re} \left[\frac{i}{\pi} \ln (-\mathbf{A} - i\omega \mathbb{I}) \right] \quad (20)$$

Proof. See A □

Now we want to rewrite the cost function (17) to a more computationally tractable form. This is done by using the realization given in (16) and by partitioning the Gramians $\mathbf{P}_{E,\omega}$ and $\mathbf{Q}_{E,\omega}$ as

$$\mathbf{P}_{E,\omega} = \begin{pmatrix} \mathbf{P}_\omega & \mathbf{X}_\omega \\ \mathbf{X}_\omega^\top & \hat{\mathbf{P}}_\omega \end{pmatrix}, \quad \mathbf{Q}_{E,\omega} = \begin{pmatrix} \mathbf{Q}_\omega & \mathbf{Y}_\omega \\ \mathbf{Y}_\omega^\top & \hat{\mathbf{Q}}_\omega \end{pmatrix}, \quad (21)$$

and $\mathbf{S}_{E,\omega}$ as

$$\mathbf{S}_{E,\omega} = \begin{pmatrix} \mathbf{S}_\omega & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{S}}_\omega \end{pmatrix}. \quad (22)$$

$\mathbf{P}_\omega, \mathbf{Q}_\omega, \hat{\mathbf{P}}_\omega, \hat{\mathbf{Q}}_\omega, \mathbf{X}_\omega$ and \mathbf{Y}_ω satisfy, due to (18), the Sylvester and Lyapunov equations

$$\mathbf{A} \mathbf{P}_\omega + \mathbf{P}_\omega \mathbf{A}^\top + \mathbf{S}_\omega \mathbf{B} \mathbf{B}^\top + \mathbf{B} \mathbf{B}^\top \mathbf{S}_\omega^* = \mathbf{0}, \quad (23a)$$

$$\mathbf{A} \mathbf{X}_\omega + \mathbf{X}_\omega \hat{\mathbf{A}}^\top + \mathbf{S}_\omega \mathbf{B} \hat{\mathbf{B}}^\top + \mathbf{B} \hat{\mathbf{B}}^\top \mathbf{S}_\omega^* = \mathbf{0}, \quad (23b)$$

$$\hat{\mathbf{A}} \hat{\mathbf{P}}_\omega + \hat{\mathbf{P}}_\omega \hat{\mathbf{A}}^\top + \hat{\mathbf{S}}_\omega \hat{\mathbf{B}} \hat{\mathbf{B}}^\top + \hat{\mathbf{B}} \hat{\mathbf{B}}^\top \hat{\mathbf{S}}_\omega^* = \mathbf{0}, \quad (23c)$$

$$\mathbf{A}^\top \mathbf{Q}_\omega + \mathbf{Q}_\omega \mathbf{A} + \mathbf{S}_\omega^* \mathbf{C}^\top \mathbf{C} + \mathbf{C}^\top \mathbf{C} \mathbf{S}_\omega = \mathbf{0}, \quad (23d)$$

$$\mathbf{A}^\top \mathbf{Y}_\omega + \mathbf{Y}_\omega \hat{\mathbf{A}} - \mathbf{S}_\omega^* \mathbf{C}^\top \hat{\mathbf{C}} - \mathbf{C}^\top \hat{\mathbf{C}} \mathbf{S}_\omega = \mathbf{0}, \quad (23e)$$

$$\hat{\mathbf{A}}^\top \hat{\mathbf{Q}}_\omega + \hat{\mathbf{Q}}_\omega \hat{\mathbf{A}} + \hat{\mathbf{S}}_\omega^* \hat{\mathbf{C}}^\top \hat{\mathbf{C}} + \hat{\mathbf{C}}^\top \hat{\mathbf{C}} \hat{\mathbf{S}}_\omega = \mathbf{0}, \quad (23f)$$

with

$$\mathbf{S}_\omega = \text{Re} \left[\frac{i}{2\pi} \ln (-\mathbf{A} - i\omega \mathbb{I}) \right], \quad \hat{\mathbf{S}}_\omega = \text{Re} \left[\frac{i}{2\pi} \ln (-\hat{\mathbf{A}} - i\omega \mathbb{I}) \right]. \quad (24)$$

Note that \mathbf{P}_ω and \mathbf{Q}_ω satisfy the Lyapunov equations for the frequency-limited controllability and observability Gramians for the given model, and $\hat{\mathbf{P}}_\omega$ and $\hat{\mathbf{Q}}_\omega$ satisfy the Lyapunov equations for the frequency-limited controllability and observability Gramians for the sought model.

With the partitioning of $\mathbf{P}_{E,\omega}$ and $\mathbf{Q}_{E,\omega}$ it is possible to rewrite (17) in two alternative forms

$$\begin{aligned} \|E\|_{\mathcal{H}_2,\omega}^2 = & \text{tr} \left(\mathbf{B}^\top \mathbf{Q}_\omega \mathbf{B} + 2\mathbf{B}^\top \mathbf{Y}_\omega \hat{\mathbf{B}} + \hat{\mathbf{B}}^\top \hat{\mathbf{Q}}_\omega \hat{\mathbf{B}} \right) \\ & + 2 \text{tr} \left[\mathbf{C} \mathbf{S}_\omega \mathbf{B} + \mathbf{D} \frac{\omega}{2\pi} - \left(\hat{\mathbf{C}} \hat{\mathbf{S}}_\omega \hat{\mathbf{B}} + \hat{\mathbf{D}} \frac{\omega}{2\pi} \right) \right] (\mathbf{D}^\top - \hat{\mathbf{D}}^\top), \end{aligned} \quad (25a)$$

$$\begin{aligned} \|E\|_{\mathcal{H}_2,\omega}^2 = & \text{tr} \left(\mathbf{C} \mathbf{P}_\omega \mathbf{C}^\top - 2\mathbf{C} \mathbf{X}_\omega \hat{\mathbf{C}}^\top + \hat{\mathbf{C}} \hat{\mathbf{P}}_\omega \hat{\mathbf{C}}^\top \right) \\ & + 2 \text{tr} \left[\mathbf{C} \mathbf{S}_\omega \mathbf{B} + \mathbf{D} \frac{\omega}{2\pi} - \left(\hat{\mathbf{C}} \hat{\mathbf{S}}_\omega \hat{\mathbf{B}} + \hat{\mathbf{D}} \frac{\omega}{2\pi} \right) \right] (\mathbf{D}^\top - \hat{\mathbf{D}}^\top). \end{aligned} \quad (25b)$$

Remark 2. Note that neither the term $\mathbf{B}^\top \mathbf{Q}_\omega \mathbf{B}$ nor the term $\mathbf{C} \mathbf{P}_\omega \mathbf{C}^\top$, which are included in the cost function (25), depend on the optimization variables, $\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}}$ and $\hat{\mathbf{D}}$. Hence, these terms can be excluded from the optimization. These are the only terms including \mathbf{P}_ω and \mathbf{Q}_ω which are the most costly to compute.

When optimizing the frequency-limited \mathcal{H}_2 -norm using the system matrices as optimization variables we have the freedom to choose which elements we optimize over, i.e., we can introduce structure in $\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}}$ and $\hat{\mathbf{D}}$, as long as we can find an $\hat{\mathbf{A}}$ that is Hurwitz. Let us introduce the matrices $\mathbf{S}_{\hat{\mathbf{A}}}, \mathbf{S}_{\hat{\mathbf{B}}}, \mathbf{S}_{\hat{\mathbf{C}}}$ and $\mathbf{S}_{\hat{\mathbf{D}}}$ which hold the structure of the sought matrices, i.e.,

$$[\mathbf{S}_{\hat{\mathbf{A}}}]_{ij} = \begin{cases} 1, & \text{if } [\hat{\mathbf{A}}]_{ij} \text{ is a free variable;} \\ 0, & \text{otherwise.} \end{cases} \quad (26)$$

We will see in the next section, that due to the element-wise differentiation, this structure will be inherited in the gradient.

The parametrization of the sought system using the full system matrices is of course redundant, which leads to a non-unique minimum of the cost function in the parameter space. This leads to a singular Hessian matrix. However, this is taken care of in most quasi-newton solvers to ensure that the minimum is reached in a numerically stable way.

3.1 Gradient of the Cost Function

An appealing feature of the proposed nonlinear optimization approach, using our proposed \mathcal{H}_2 -like measure to solve the problem, is that the equations (25) are differentiable in the system matrices, $\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}}$ and $\hat{\mathbf{D}}$. In addition, the closed form expression obtained when differentiating the cost function is expressed in the given data ($\mathbf{A}, \mathbf{B}, \mathbf{C}$ and \mathbf{D}), the optimization variables ($\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}}$ and $\hat{\mathbf{D}}$) and solutions to the equations in (23).

To show this we start by differentiating with respect to $\hat{\mathbf{B}}, \hat{\mathbf{C}}$ and $\hat{\mathbf{D}}$. First we note that neither $\mathbf{Q}_\omega, \mathbf{Y}_\omega$ nor $\hat{\mathbf{Q}}_\omega$ in equation (25a) depends on $\hat{\mathbf{B}}$ which means that the equation is quadratic in $\hat{\mathbf{B}}$. Analogous observations can be made with equation (25b) and the variable $\hat{\mathbf{C}}$ and similarly with $\hat{\mathbf{D}}$. Hence, the derivative

of the cost function with respect $\hat{\mathbf{B}}, \hat{\mathbf{C}}$ and $\hat{\mathbf{D}}$ becomes

$$\frac{\partial \|E\|_{\mathcal{H}_2, \omega}^2}{\partial \hat{\mathbf{B}}} = 2 \left(\hat{\mathbf{Q}}_\omega \hat{\mathbf{B}} + \mathbf{Y}_\omega^\top \mathbf{B} - \hat{\mathbf{S}}_\omega^\top \hat{\mathbf{C}}^\top (\mathbf{D} - \hat{\mathbf{D}}) \right) \odot \mathbf{S}_{\hat{\mathbf{B}}}, \quad (27a)$$

$$\frac{\partial \|E\|_{\mathcal{H}_2, \omega}^2}{\partial \hat{\mathbf{C}}} = 2 \left(\hat{\mathbf{C}} \hat{\mathbf{P}}_\omega - \mathbf{C} \mathbf{X}_\omega - (\mathbf{D} - \hat{\mathbf{D}}) \hat{\mathbf{B}}^\top \hat{\mathbf{S}}_\omega^\top \right) \odot \mathbf{S}_{\hat{\mathbf{C}}}, \quad (27b)$$

$$\frac{\partial \|E\|_{\mathcal{H}_2, \omega}^2}{\partial \hat{\mathbf{D}}} = -2 \left(\mathbf{C} \mathbf{S}_\omega \mathbf{B} + \mathbf{D} \frac{\omega}{2\pi} - \hat{\mathbf{C}} \hat{\mathbf{S}}_\omega \hat{\mathbf{B}} - \hat{\mathbf{D}} \frac{\omega}{2\pi} + (\mathbf{D} - \hat{\mathbf{D}}) \frac{\omega}{2\pi} \right) \odot \mathbf{S}_{\hat{\mathbf{D}}}, \quad (27c)$$

where \odot represents the Hadamard product of matrices, i.e., element-wise multiplication.

For the more complicated case of differentiating with respect to $\hat{\mathbf{A}}$ we observe that $\hat{\mathbf{Q}}_\omega$ and \mathbf{Y}_ω do depend on $\hat{\mathbf{A}}$, see the equations in (23). The calculations of this part of the gradient are lengthy and can be found in B.

The complete gradient becomes

$$\frac{\partial \|E\|_{\mathcal{H}_2, \omega}^2}{\partial \hat{\mathbf{A}}} = 2 \left(\mathbf{Y}_\omega^\top \mathbf{X} + \hat{\mathbf{Q}}_\omega \hat{\mathbf{P}} \right) \odot \mathbf{S}_{\hat{\mathbf{A}}} - 2 \mathbf{W} \odot \mathbf{S}_{\hat{\mathbf{A}}}, \quad (28a)$$

$$\frac{\partial \|E\|_{\mathcal{H}_2, \omega}^2}{\partial \hat{\mathbf{B}}} = 2 \left(\hat{\mathbf{Q}}_\omega \hat{\mathbf{B}} + \mathbf{Y}_\omega^\top \mathbf{B} \hat{\mathbf{S}}_\omega^\top \hat{\mathbf{C}}^\top (\mathbf{D} - \hat{\mathbf{D}}) \right) \odot \mathbf{S}_{\hat{\mathbf{B}}}, \quad (28b)$$

$$\frac{\partial \|E\|_{\mathcal{H}_2, \omega}^2}{\partial \hat{\mathbf{C}}} = 2 \left(\hat{\mathbf{C}} \hat{\mathbf{P}}_\omega - \mathbf{C} \mathbf{X}_\omega - (\mathbf{D} - \hat{\mathbf{D}}) \hat{\mathbf{B}}^\top \hat{\mathbf{S}}_\omega^\top \right) \odot \mathbf{S}_{\hat{\mathbf{C}}}, \quad (28c)$$

$$\frac{\partial \|E\|_{\mathcal{H}_2, \omega}^2}{\partial \hat{\mathbf{D}}} = -2 \left(\mathbf{C} \mathbf{S}_\omega \mathbf{B} + \mathbf{D} \omega - \hat{\mathbf{C}} \hat{\mathbf{S}}_\omega \hat{\mathbf{B}} - \hat{\mathbf{D}} \omega + (\mathbf{D} - \hat{\mathbf{D}}) \omega \right) \odot \mathbf{S}_{\hat{\mathbf{D}}}, \quad (28d)$$

where

$$\mathbf{W} = \text{Re} \left[\frac{i}{\pi} L \left(-\hat{\mathbf{A}} - i\omega \mathbb{I}, \mathbf{V} \right) \right]^\top, \quad (29a)$$

$$\mathbf{V} = \hat{\mathbf{C}}^\top \hat{\mathbf{C}} \hat{\mathbf{P}} - \hat{\mathbf{C}}^\top \mathbf{C} \mathbf{X} - \hat{\mathbf{C}}^\top (\mathbf{D} - \hat{\mathbf{D}}) \hat{\mathbf{B}}^\top \quad (29b)$$

with the function $L(\cdot, \cdot)$ being the Frechét derivative of the matrix logarithm, see [5].

Remark 3. Remember that if $\omega = \infty$ is included as the end frequency we need to have $\mathbf{D}_E = \mathbf{D} - \hat{\mathbf{D}} = \mathbf{0}$, i.e., $\hat{\mathbf{D}}$ constant and with $\mathbf{D} = \hat{\mathbf{D}}$.

Remark 4. The proposed method can, analogous to what is done in [7], easily be extended to a method for identifying LPV-models over a limited frequency domain.

Remark 5. By supplying the cost function and its gradient in computationally efficient forms this method can be used in any off-the-shelf Quasi-Newton solver.

Remark 6. By using a stable model, e.g., a model from a Hankel reduction, as an initial point in the optimization and using a line-search we can limit the search to stable models.

4 Numerical Examples

In this section three examples are used to illustrate the applicability of the method and to compare it with other methods. In the examples we will use three different methods; Truncation of Hankel singular values (will be called Hankel), the method proposed in [2] (called Gawronski), the method proposed in [4] (called Mod. Gawronski) and the proposed method (called Prop. method). The proposed method and the original and modified Gawronski method will take the limited frequency range into account, but Hankel will not. We will also compare with the Hankel method where we use an input filter to help that method to focus on the frequency interval of interest. The Gawronski method is a representative method among frequency-weighted methods, see [4], with the benefit of not having to design weighting functions, but with the drawback that it cannot guarantee that the resulting model is stable.

The proposed method uses a cost function which is non-convex, which makes it important to use a good initial point. For the examples presented here we have used the model obtained by the Gawronski method as an initial point for the optimization in the proposed method. If the Gawronski method generates an unstable model we use the Hankel method instead.

To evaluate the different models against each other we will compare the error model, $G - \hat{G}$, for the given frequency interval using the limited frequency \mathcal{H}_2 -norm, $\|G - \hat{G}\|_{\mathcal{H}_2, \omega}$, and the relative limited frequency \mathcal{H}_2 -norm, $\frac{\|G - \hat{G}\|_{\mathcal{H}_2, \omega}}{\|G\|_{\mathcal{H}_2, \omega}}$. We will also compute the relative \mathcal{H}_∞ -norm on the given frequency interval, denoted $\frac{\|G - \hat{G}\|_{\mathcal{H}_\infty, \omega}}{\|G\|_{\mathcal{H}_\infty, \omega}}$, and the eigenvalue with the largest real part.

Example 1 (Small illustrative example). *This example addresses a small model with four states. The model is composed of two second order models in series, one with a resonance frequency at $\omega = 1$ and the other at $\omega = 3$. We will limit the frequency range to $\omega \in [0, 1.7]$ to try to only capture the first model.*

$$G = G_1 G_2 = \frac{1}{s^2 + 0.2s + 1} \frac{9}{s^2 + 0.003s + 9}. \quad (30)$$

To help the Hankel method we create a low pass Butterworth filter of order 10 with a cut off frequency of 1.7, see Figure 1. The results from the different methods can be seen in Figures 2 and 3 and Table 1. As can be seen in the result we are successful in finding a good model for the first model with both the proposed method and the Gawronski method. All the reduced order models are stable. The Hankel method and the modified Gawronski method captures the wrong resonance mode (from our perspective) and fails completely in the lower frequency region. The Hankel method on the filtered model finds a model with the same resonance frequency, but otherwise has a bad correspondence with the true model. The proposed method and the Gawronski method return models essentially indiscernible.

Example 2 (Example 1 in [2]). *In this example we reuse Example 1 from [2]. The model, which is a spring-damper model with three masses, has six states and we will reduce the model to three states and limit the frequency interval to $\omega \in [1.5, 3.2]$. To help the Hankel method we create a band pass Butterworth filter of order 10 with cut off frequencies of 1.5 and 3.2 rad/s, see Figure 4.*

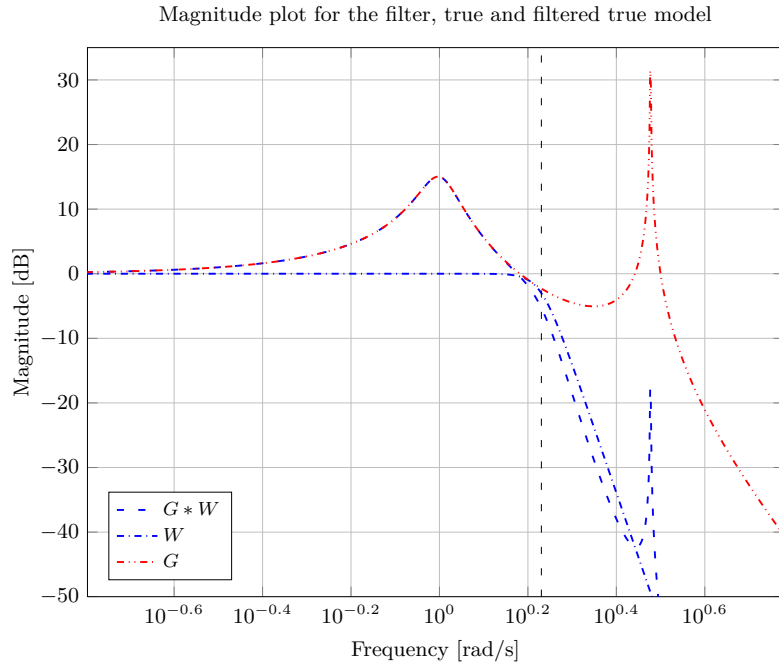


Figure 1: Magnitude plot of the given model, the filtered model and the filter used in Example 1. For a more easy read plot, in color, the reader is referred to the digital version of this paper.

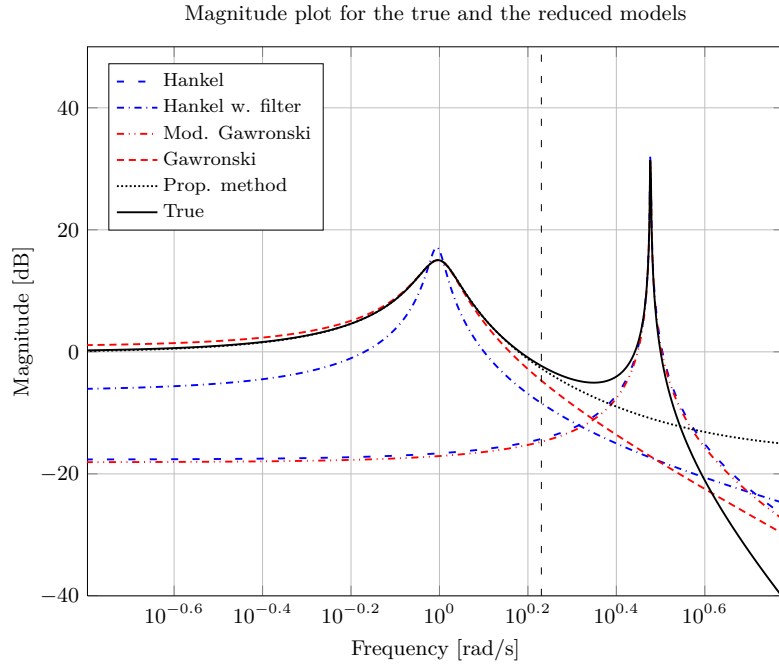


Figure 2: Magnitude plot of the given and reduced order models in Example 1 for $\omega \in [0, 1.7]$. The dashed black vertical line denotes $\omega = 1.7$. We see in the figure that the proposed method and the Gawronski method finds models which are good approximations of the true model on the given frequency interval. For a more easy read plot, in color, the reader is referred to the digital version of this paper.

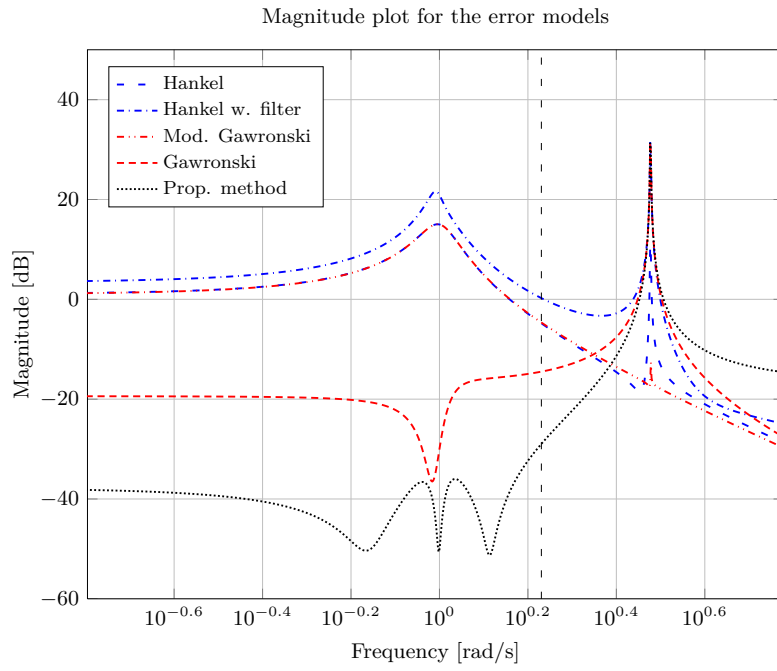


Figure 3: Magnitude plots of the error models in Example 1 for $\omega \in [0, 1.7]$. The dashed black vertical line denotes $\omega = 1.7$. We see in the figure that the proposed method and the Gawronski method finds accurate approximations to the true model on the given frequency interval. For a more easy read plot, in color, the reader is referred to the digital version of this paper.

Table 1: Numerical results for Example 1

	$\ G - \hat{G}\ _{\mathcal{H}_2, \omega}$	$\frac{\ G - \hat{G}\ _{\mathcal{H}_2, \omega}}{\ G\ _{\mathcal{H}_2, \omega}}$	$\frac{\ G - \hat{G}\ _{\mathcal{H}_\infty, \omega}}{\ G\ _{\mathcal{H}_\infty, \omega}}$	$\text{Re } \lambda_{\max}$
Hankel	1.77e+00	1.01e+00	1.00e+00	-1.59e-03
Hankel w. filter	2.93e+00	1.67e+00	2.14e+00	-4.03e-02
Mod. Gawronski	1.77e+00	1.01e+00	1.00e+00	-1.51e-03
Gawronski	9.14e-02	5.21e-02	3.35e-02	-9.88e-02
Prop. method	8.51e-02	4.85e-02	3.26e-02	-9.94e-02

Magnitude plot for the filter, true and filtered true model

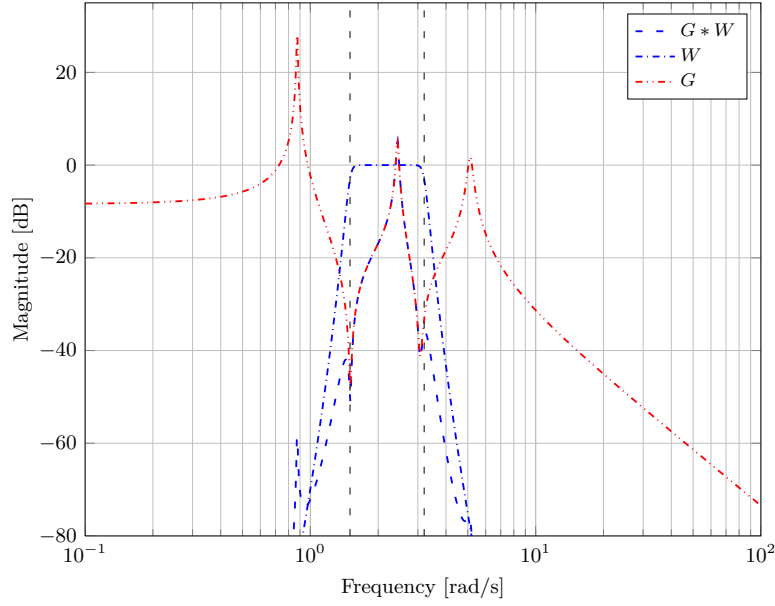


Figure 4: Magnitude plot of the given model, the filtered model and the filter used in Example 2. For a more easy read plot, in color, the reader is referred to the digital version of this paper.

The results from the different methods can be seen in Figures 5 and 6 and Table 2. The proposed method and the Gawronski method are also in this example successful in finding low order models that approximate the given model on the given frequency range, and all the reduced order models are stable. The proposed method and the Gawronski method captures the correct frequency interval with good accuracy. The Hankel method on the filtered model captures the correct resonance frequency, however, with the worst overall \mathcal{H}_2 -measure. The other two methods misses the relevant frequency interval. Only the proposed method, the Gawronski method and the Hankel method with the filtered model are able to find the right resonance peak, whilst the modified Gawronski and the Hankel method are completely off.

Example 3 (Aircraft example). The model in this example is a model with 22 states that describes the longitudinal motion of an aircraft, see [9]. We will

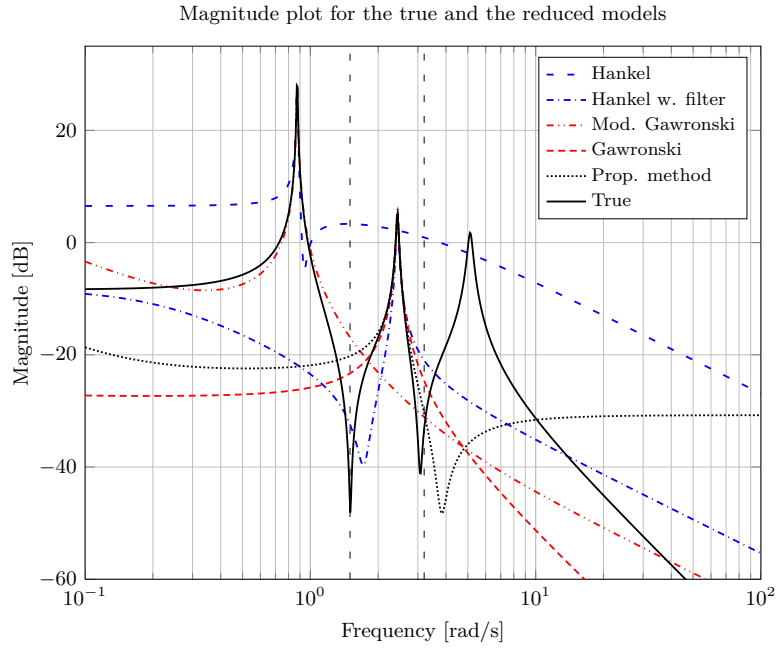


Figure 5: Magnitude plot of the given and reduced order models in Example 2 for $\omega \in [1.5, 3.2]$. The dashed black vertical lines denotes $\omega = 1.5$ and 3.2 . We see that proposed method and the Gawronski method tries to capture the correct resonance frequency with good accuracy, also the Hankel method on the filtered model does this but not as good. The other methods misses the relevant frequency interval. For a more easy read plot, in color, the reader is referred to the digital version of this paper.

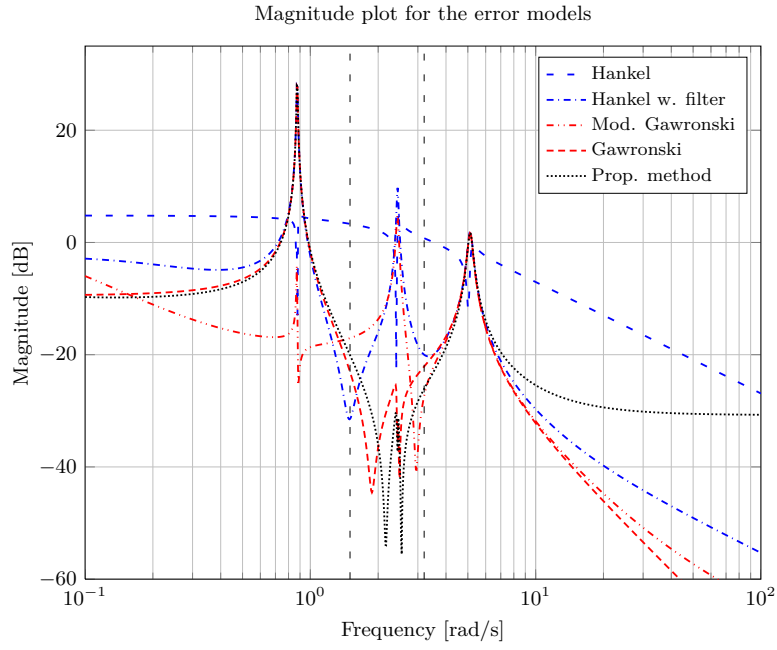


Figure 6: Magnitude plots of the error models in Example 2 for $\omega \in [1.5, 3.2]$. The dashed black vertical lines denotes $\omega = 1.5$ and 3.2 . The proposed method and the Gawronski method finds the best approximations on the relevant frequency interval. For a more easy read plot, in color, the reader is referred to the digital version of this paper.

Table 2: Numerical results for Example 2

	$\ G - \hat{G}\ _{\mathcal{H}_{2,\omega}}$	$\frac{\ G - \hat{G}\ _{\mathcal{H}_{2,\omega}}}{\ G\ _{\mathcal{H}_{2,\omega}}}$	$\frac{\ G - \hat{G}\ _{\mathcal{H}_{\infty,\omega}}}{\ G\ _{\mathcal{H}_{\infty,\omega}}}$	$\text{Re } \lambda_{\max}$
Hankel	9.27e-01	3.05e+00	9.59e-01	-3.54e-03
Hankel w. filter	4.58e-01	1.50e+00	1.71e+00	-2.02e-02
Mod. Gawronski	3.10e-01	1.02e+00	9.89e-01	-3.78e-03
Gawronski	3.30e-02	1.09e-01	4.36e-02	-2.86e-02
Prop. method	2.72e-02	8.94e-02	5.34e-02	-2.86e-02

Magnitude plot for the filter, true and filtered true model

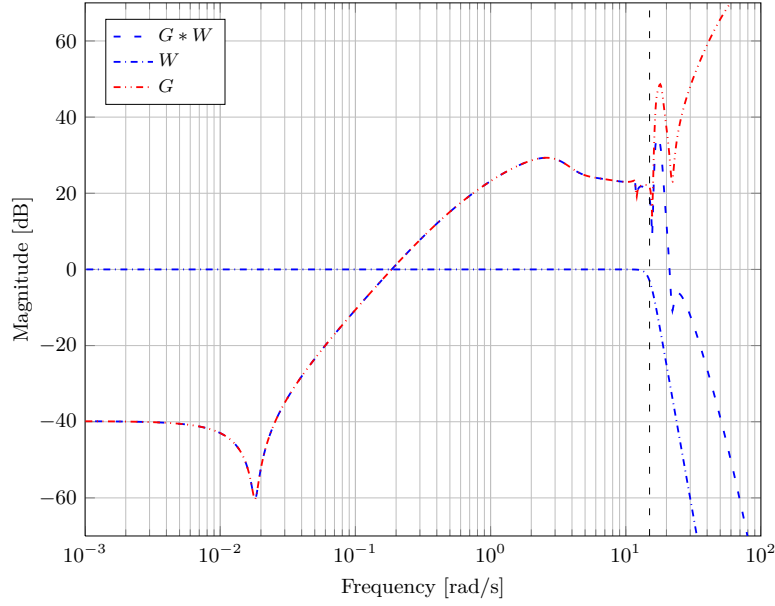


Figure 7: Magnitude plot of the given model, the filtered model and the filter used in Example 3. For a more easy read plot, in color, the reader is referred to the digital version of this paper.

reduce this model to 15 states and limit the frequency range to $\omega \in [0, 15]$. To help the Hankel method we create a low pass Butterworth filter of order 10 with a cut off frequency of 15 rad/s, see Figure 7. The results from the different methods can be seen in Figures 8 and 9 and Table 3. In this example the Gawronski method results in a model which at a first glance looks very good, however the model is unstable. The other methods capture the given model with varying accuracy. Since the Gawronski method generates an unstable model in this example, the proposed method is initialized with the model from the Hankel reduction instead. The proposed method finds the most accurate model on the given frequency range.

In all three of the above examples we observe that the proposed method finds an at least as good model as the method proposed in [2], but with the proposed method we can guarantee that the reduced order model is stable and

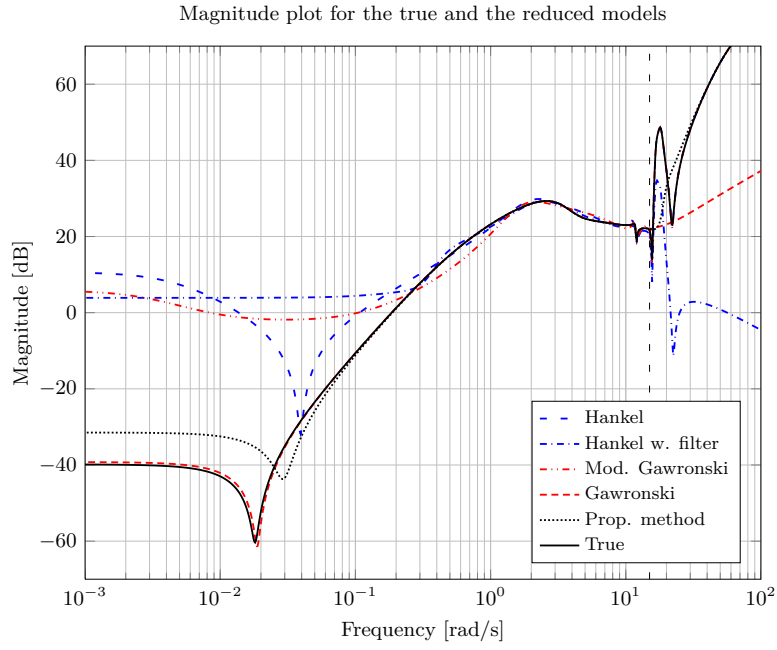


Figure 8: Magnitude plot of the given and reduced order models in Example 3 for $\omega \in [0, 15]$. The dashed black vertical line denotes $\omega = 15$. The Gawronski method looks to have found the best approximation, however the model found by the Gawronski method is unstable. The proposed method finds the best stable approximation of the true model. For a more easy read plot, in color, the reader is referred to the digital version of this paper.

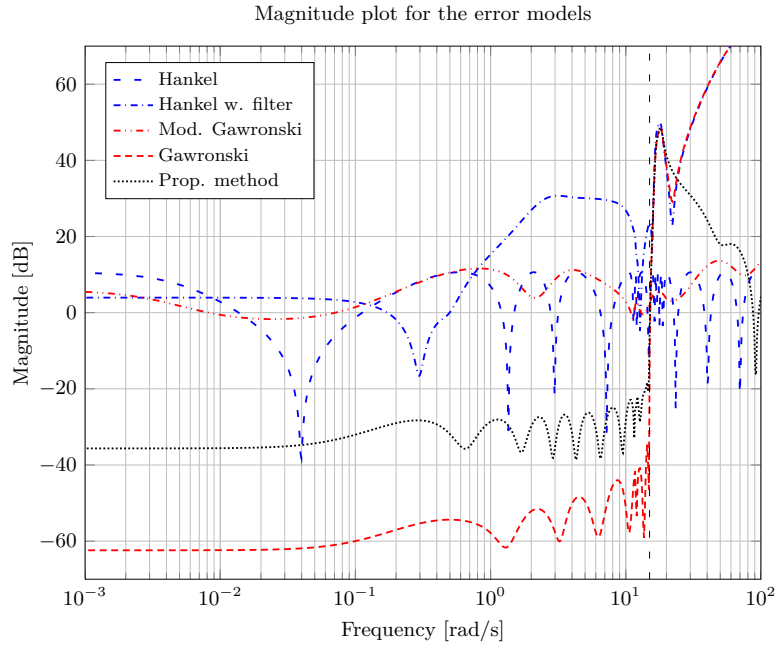


Figure 9: Magnitude plots of the error models in Example 3 for $\omega \in [0, 15]$. The dashed black vertical line denotes $\omega = 15$. The Gawronski method looks to have found the approximation with the smallest error, however the model found by the Gawronski method is unstable. The proposed method finds the model with the lowest error which is stable. For a more easy read plot, in color, the reader is referred to the digital version of this paper.

Table 3: Numerical results for Example 3

	$\ G - \hat{G}\ _{\mathcal{H}_2, \omega}$	$\frac{\ G - \hat{G}\ _{\mathcal{H}_2, \omega}}{\ G\ _{\mathcal{H}_2, \omega}}$	$\frac{\ G - \hat{G}\ _{\mathcal{H}_\infty, \omega}}{\ G\ _{\mathcal{H}_\infty, \omega}}$	Re λ_{\max}
Hankel	5.15e+00	1.39e-01	1.16e-01	-4.85e-03
Hankel w. filter	5.10e+01	1.37e+00	1.17e+00	-1.24e-01
Mod. Gawronski	5.00e+00	1.35e-01	1.30e-01	-3.19e-03
Gawronski	—	—	—	4.67e+00
Prop. method	1.68e-01	4.51e-03	1.51e-02	-1.77e-01

also impose structure in the system matrices. In the first two examples it takes less than one second and in the third example about 20 seconds for the proposed method to find a reduced order model.

5 Conclusions

In this paper we have proposed a method, based on nonlinear optimization, that uses the frequency-limited Gramians introduced in [2], and the method does not have the drawback of finding unstable models. We use these Gramians to construct a frequency-limited \mathcal{H}_2 -norm which describes the cost function of the optimization problem. We derive a gradient of the proposed cost function which enables us to use off-the-shelf optimization software to solve the problem efficiently. The derivation of the method also enables us to impose structural constraints, e.g., upper triangular \mathbf{A} -matrix, in the system matrices. The derivation follows closely the technique in [7] and it is easy to extend the method to identify LPV-models. We also presented three examples of different sizes and characteristics to show the applicability of the method. The method uses a nonlinear optimization approach on a nonconvex problem, hence the unavoidable problem of local minima are present. However, we consider this as a two step approach where you use your favorite reduction method as an initial point to this method which will refine the solution, as we have shown in the examples.

References

- [1] Dale F. Enns. Model reduction with balanced realizations: An error bound and a frequency weighted generalization. In *Proceedings of the 23rd IEEE Conference on Decision and Control*, pages 127 – 132, Las Vegas, USA, 1984.
- [2] Wodek Gawronski and Jer-Nan Juang. Model reduction in limited time and frequency intervals. *International Journal of Systems Science*, 21(2):349–376, 1990.
- [3] Abdul Ghafoor and Victor Sreeram. A survey/review of frequency-weighted balanced model reduction techniques. *Journal of Dynamic Systems, Measurement and Control*, 130:061004, 2008.

- [4] Serkan Gugercin and Athanasios C. Antoulas. A survey of model reduction by balanced truncation and some new results. *International Journal of Control*, 77(8):748–766, 2004.
- [5] Nicholas J. Higham. *Functions of Matrices: Theory and Computation*. SIAM, 2008.
- [6] Ching-An Lin and Tai-Yih Chiu. Model reduction via frequency weighted balanced realization. *Control Theory and Advanced Technology*, 8:341–351, 1992.
- [7] Daniel Petersson. *Nonlinear optimization approaches to \mathcal{H}_2 -norm based LPV modelling and control*. Licentiate thesis no. 1453, Department of Electrical Engineering, Linköping University, 2010.
- [8] Shafishuhaza Sahlan, Abdul Ghafoor, and Victor Sreeram. A new method for the model reduction technique via a limited frequency interval impulse response gramian. *Mathematical and Computer Modelling*, 55(3-4):1034–1040, 2012.
- [9] Andreas Varga, Anders Hansson, and Guilhem Puyou, editors. *Optimization Based Clearance of Flight Control Laws*. Lecture Notes in Control and Information Science. Springer, 2012.
- [10] G. Wang, Victor Sreeram, and W. Q. Liu. A new frequency-weighted balanced truncation method and an error bound. *IEEE Trans. Autom. Control*, 44(9):1734 – 1737, 1999.
- [11] Kemin Zhou, John C. Doyle, and Keith Glover. *Robust and optimal control*. Prentice-Hall, Inc., 1996.

A Proof of Lemma 1

We want to show that

$$\mathbf{S}_\omega = \frac{i}{2\pi} \ln ((\mathbf{A} + i\omega\mathbb{I})(\mathbf{A} - i\omega\mathbb{I})^{-1}) = \text{Re} \left[\frac{i}{\pi} \ln (-\mathbf{A} - i\omega\mathbb{I}) \right] \quad (31)$$

assuming that \mathbf{A} is Hurwitz, i.e., \mathbf{A} is a real matrix and if λ is an eigenvalue to \mathbf{A} then we have that $\text{Re}(\lambda) < 0$. We can decompose \mathbf{A} as $\mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}$ where \mathbf{D} is a diagonal matrix with the eigenvalues to \mathbf{A} on the diagonal and \mathbf{V} is the matrix with the eigenvectors to \mathbf{A} as its columns. For a real matrix we have that the eigenvectors for real eigenvalues are always real and that the eigenvectors for a complex-conjugate pair are also complex-conjugate. We can write \mathbf{A} as

$$\begin{aligned} \mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1} &= [v_1 v_2 \cdots v_n] \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} \tilde{v}_1^T \\ \tilde{v}_2^T \\ \vdots \\ \tilde{v}_n^T \end{bmatrix} \\ &= v_1 \tilde{v}_1 \lambda_1 + v_2 \tilde{v}_2 \lambda_2 + \cdots + v_n \tilde{v}_n \lambda_n \end{aligned} \quad (32)$$

and $f(\mathbf{A})$ as

$$\begin{aligned} f(\mathbf{A}) &= \mathbf{V}f(\mathbf{D})\mathbf{V}^{-1} = [v_1 v_2 \cdots v_n] \begin{bmatrix} f(\lambda_1) & 0 & \cdots & 0 \\ 0 & f(\lambda_2) & & \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & f(\lambda_n) \end{bmatrix} \begin{bmatrix} \tilde{v}_1^\top \\ \tilde{v}_2^\top \\ \vdots \\ \tilde{v}_n^\top \end{bmatrix} \\ &= v_1 \tilde{v}_1 f(\lambda_1) + v_2 \tilde{v}_2 f(\lambda_2) + \cdots + v_n \tilde{v}_n f(\lambda_n). \end{aligned} \quad (33)$$

For a real eigenvalue, λ , in \mathbf{A} with corresponding eigenvector, v , represented as one term in equation (33) and

$$f(\lambda) = i \ln \left(\frac{\lambda + i\omega}{\lambda - i\omega} \right) = i \ln(-\lambda - i\omega) - i \ln(-\lambda + i\omega) \quad (34)$$

we can write this term as

$$\begin{aligned} v \tilde{v}^\top (i \ln(-\lambda - i\omega) - i \ln(-\lambda + i\omega)) &= v \tilde{v}^\top \left(i \ln(-\lambda - i\omega) + \overline{i \ln(-\lambda - i\omega)} \right) \\ &= 2v \tilde{v}^\top \operatorname{Re} [i \ln(-\lambda - i\omega)] = 2 \operatorname{Re} [v \tilde{v}^\top i \ln(-\lambda - i\omega)] = 2 \operatorname{Re} [v \tilde{v}^\top g(\lambda)]. \end{aligned} \quad (35)$$

If we now instead look at two terms in (33) corresponding to a complex-conjugate pair of eigenvalues, λ and $\bar{\lambda}$, with corresponding eigenvectors, v and \bar{v} , then we can write this as

$$\begin{aligned} v \tilde{v}^\top f(\lambda) + \bar{v} \bar{\tilde{v}}^\top f(\bar{\lambda}) &= v \tilde{v}^\top i \ln(-\lambda - i\omega) - v \tilde{v}^\top i \ln(-\lambda + i\omega) \\ &\quad + \bar{v} \bar{\tilde{v}}^\top i \ln(-\bar{\lambda} - i\omega) - \bar{v} \bar{\tilde{v}}^\top i \ln(-\bar{\lambda} + i\omega) = v \tilde{v}^\top g(\lambda) + \overline{v \tilde{v}^\top g(\lambda)} + \\ &\quad + \bar{v} \bar{\tilde{v}}^\top g(\bar{\lambda}) + \overline{\bar{v} \bar{\tilde{v}}^\top g(\bar{\lambda})} = 2 \operatorname{Re} [v \tilde{v}^\top g(\lambda)] + 2 \operatorname{Re} [\bar{v} \bar{\tilde{v}}^\top g(\bar{\lambda})]. \end{aligned} \quad (36)$$

This means that for a matrix, \mathbf{A} , which is Hurwitz, we have

$$\begin{aligned} f(\mathbf{A}) &= \mathbf{V}f(\mathbf{D})\mathbf{V}^{-1} = v_1 \tilde{v}_1 f(\lambda_1) + v_2 \tilde{v}_2 f(\lambda_2) + \cdots + v_n \tilde{v}_n f(\lambda_n) \\ &= 2 \operatorname{Re} [v_1 \tilde{v}_1 g(\lambda_1) + v_2 \tilde{v}_2 g(\lambda_2) + \cdots + v_n \tilde{v}_n g(\lambda_n)] = 2 \operatorname{Re} g(\mathbf{A}), \end{aligned} \quad (37)$$

i.e., we have

$$\mathbf{S}_\omega = \frac{i}{2\pi} \ln((\mathbf{A} + i\omega\mathbb{I})(\mathbf{A} - i\omega\mathbb{I})^{-1}) = \operatorname{Re} \left[\frac{i}{\pi} \ln(-\mathbf{A} - i\omega\mathbb{I}) \right]. \quad (38)$$

B Derivation of the Gradient w.r.t. $\hat{\mathbf{A}}$

In this appendix we present the differentiation of the cost function with respect to $\hat{\mathbf{A}}$.

The cost function is

$$\begin{aligned} ||E||_{\mathcal{H}_{2,\omega}}^2 &= \operatorname{tr} \left(\mathbf{B}^\top \mathbf{Q}_\omega \mathbf{B} + 2\mathbf{B}^\top \mathbf{Y}_\omega \hat{\mathbf{B}} + \hat{\mathbf{B}}^\top \hat{\mathbf{Q}}_\omega \hat{\mathbf{B}} \right) \\ &\quad + 2 \operatorname{tr} \left[\mathbf{C} \mathbf{S}_\omega \mathbf{B} + \mathbf{D} \frac{\omega}{2\pi} - \left(\hat{\mathbf{C}} \hat{\mathbf{S}}_\omega \hat{\mathbf{B}} + \hat{\mathbf{D}} \frac{\omega}{2\pi} \right) \right] (\mathbf{D}^\top - \hat{\mathbf{D}}^\top) \end{aligned} \quad (39)$$

and by looking at the equations

$$\mathbf{A}^\top \mathbf{Y}_\omega + \mathbf{Y}_\omega \hat{\mathbf{A}} - \mathbf{S}_\omega^* \mathbf{C}^\top \hat{\mathbf{C}} - \mathbf{C}^\top \hat{\mathbf{C}} \hat{\mathbf{S}}_\omega = \mathbf{0}, \quad (40a)$$

$$\hat{\mathbf{A}}^\top \hat{\mathbf{Q}}_\omega + \hat{\mathbf{Q}}_\omega \hat{\mathbf{A}} + \hat{\mathbf{S}}_\omega^* \hat{\mathbf{C}}^\top \hat{\mathbf{C}} + \hat{\mathbf{C}}^\top \hat{\mathbf{C}} \hat{\mathbf{S}}_\omega = \mathbf{0}, \quad (40b)$$

we observe that $\hat{\mathbf{Q}}_\omega$ and \mathbf{Y}_ω depend on $\hat{\mathbf{A}}$ which we need to keep in mind when differentiating (39) with respect to $\hat{\mathbf{A}}$. Hence, $\left[\frac{\partial \|E\|_{\mathcal{H}_{2,\omega}}^2}{\partial \hat{\mathbf{A}}} \right]_{ij}$ becomes

$$\left[\frac{\partial \|E\|_{\mathcal{H}_{2,\omega}}^2}{\partial \hat{\mathbf{A}}} \right]_{ij} = \text{tr} \left(2\hat{\mathbf{B}}\hat{\mathbf{B}}^\top \frac{\partial \mathbf{Y}_\omega}{\partial \hat{a}_{ij}} + \hat{\mathbf{B}}\hat{\mathbf{B}}^\top \frac{\partial \hat{\mathbf{Q}}_\omega}{\partial \hat{a}_{ij}} - 2\hat{\mathbf{C}} \frac{\partial \hat{\mathbf{S}}_\omega}{\partial \hat{a}_{ij}} \hat{\mathbf{B}} (\mathbf{D}^\top - \hat{\mathbf{D}}^\top) \right), \quad (41)$$

where $\frac{\partial \mathbf{Y}_\omega}{\partial \hat{a}_{ij}}$ and $\frac{\partial \hat{\mathbf{Q}}_\omega}{\partial \hat{a}_{ij}}$ depend on $\hat{\mathbf{A}}$ via the differentiated versions of the equations in (40),

$$\hat{\mathbf{A}}^\top \frac{\partial \mathbf{Y}_\omega}{\partial \hat{a}_{ij}} + \frac{\partial \mathbf{Y}_\omega}{\partial \hat{a}_{ij}} \mathbf{A} + \frac{\partial \hat{\mathbf{A}}^\top}{\partial \hat{a}_{ij}} \mathbf{Y}_\omega - \frac{\partial \hat{\mathbf{S}}_\omega^*}{\partial \hat{a}_{ij}} \hat{\mathbf{C}}^\top \mathbf{C} = \mathbf{0}, \quad (42a)$$

$$\hat{\mathbf{A}}^\top \frac{\partial \hat{\mathbf{Q}}_\omega}{\partial \hat{a}_{ij}} + \frac{\partial \hat{\mathbf{Q}}_\omega}{\partial \hat{a}_{ij}} \hat{\mathbf{A}} + \frac{\partial \hat{\mathbf{A}}^\top}{\partial \hat{a}_{ij}} \hat{\mathbf{Q}}_\omega + \hat{\mathbf{Q}}_\omega \frac{\partial \hat{\mathbf{A}}}{\partial \hat{a}_{ij}} + \frac{\partial \hat{\mathbf{S}}_\omega^*}{\partial \hat{a}_{ij}} \hat{\mathbf{C}}^\top \hat{\mathbf{C}} + \hat{\mathbf{C}}^\top \hat{\mathbf{C}} \frac{\partial \hat{\mathbf{S}}_\omega}{\partial \hat{a}_{ij}} = \mathbf{0}. \quad (42b)$$

To be able to substitute $\frac{\partial \hat{\mathbf{Q}}_\omega}{\partial \hat{a}_{ij}}$ and $\frac{\partial \mathbf{Y}_\omega}{\partial \hat{a}_{ij}}$ to something more computationally tractable we use the following lemma.

Lemma 2. *If \mathbf{M} and \mathbf{N} satisfy the Sylvester equations*

$$\mathbf{A}\mathbf{M} + \mathbf{M}\mathbf{B} + \mathbf{C} = \mathbf{0}, \quad \mathbf{N}\mathbf{A} + \mathbf{B}\mathbf{N} + \mathbf{D} = \mathbf{0},$$

then $\text{tr } \mathbf{C}\mathbf{N} = \text{tr } \mathbf{D}\mathbf{M}$.

Studying Lemma 2, the two factors in front of $\frac{\partial \mathbf{Y}_\omega}{\partial \hat{a}_{ij}}$ and $\frac{\partial \hat{\mathbf{Q}}_\omega}{\partial \hat{a}_{ij}}$ in (41) and the structure of the Lyapunov/Sylvester equations in (42), $\hat{\mathbf{A}}^\top \cdot + \cdot \hat{\mathbf{A}} + \star = 0$ and $\hat{\mathbf{A}}^\top \cdot + \cdot \mathbf{A} + \star = 0$, brings us to the conclusion that, to do the substitution, we need to solve two additional Lyapunov/Sylvester equations, namely

$$\mathbf{A}\mathbf{X} + \mathbf{X}\hat{\mathbf{A}}^\top + \hat{\mathbf{B}}\hat{\mathbf{B}}^\top = \mathbf{0}, \quad (43a)$$

$$\hat{\mathbf{A}}\hat{\mathbf{P}} + \hat{\mathbf{P}}\mathbf{A}^\top + \hat{\mathbf{B}}\hat{\mathbf{B}}^\top = \mathbf{0}. \quad (43b)$$

Note that $\hat{\mathbf{P}}$ in (43b) is the controllability Gramian for the reduced order model.

Rewriting (41) using Lemma 2, (42) and (43) leads to

$$\begin{aligned} \left[\frac{\partial \|E\|_{\mathcal{H}_{2,\omega}}^2}{\partial \hat{\mathbf{A}}} \right]_{ij} &= 2 \text{tr} \left[\frac{\partial \hat{\mathbf{A}}^\top}{\partial \hat{a}_{ij}} (\mathbf{Y}_\omega^\top \mathbf{X} + \hat{\mathbf{Q}}_\omega \hat{\mathbf{P}}) + \frac{\partial \hat{\mathbf{S}}_\omega^*}{\partial \hat{a}_{ij}} (\hat{\mathbf{C}}^\top \hat{\mathbf{C}} \hat{\mathbf{P}} - \hat{\mathbf{C}}^\top \mathbf{C} \mathbf{X}) \right] \\ &\quad - 2 \text{tr} \left[\frac{\partial \hat{\mathbf{S}}_\omega}{\partial \hat{a}_{ij}} [\hat{\mathbf{B}} (\mathbf{D}^\top - \hat{\mathbf{D}}^\top) \hat{\mathbf{C}}] \right]. \end{aligned} \quad (44)$$

What remains is to rewrite the two last terms in (44), which includes $\frac{\partial \hat{\mathbf{S}}_\omega}{\partial \hat{a}_{ij}}$ and $\frac{\partial \hat{\mathbf{S}}_\omega^*}{\partial \hat{a}_{ij}}$. Recall the definition of $\hat{\mathbf{S}}_\omega$,

$$\hat{\mathbf{S}}_\omega = \text{Re} \left[\frac{i}{\pi} \ln (-\hat{\mathbf{A}} - i\omega \mathbb{I}) \right] = \text{Re} \left[\frac{i}{\pi} \ln (r(\hat{\mathbf{A}})) \right] \quad (45)$$

and differentiate with respect to an element in $\hat{\mathbf{A}}$, i.e., a_{ij} . This yields

$$\frac{\partial \hat{\mathbf{S}}_\omega}{\partial a_{ij}} = \text{Re} \left[\frac{i}{2\pi} L \left(r(\hat{\mathbf{A}}), \frac{\partial r(\hat{\mathbf{A}})}{\partial a_{ij}} \right) \right] = \text{Re} \left[\frac{i}{2\pi} L \left(r(\hat{\mathbf{A}}), -\frac{\partial \hat{\mathbf{A}}}{\partial a_{ij}} \right) \right] \quad (46)$$

where $L(\mathbf{A}, \mathbf{E})$ is the Frechét derivative of the matrix logarithm, see [5], with

$$L(\mathbf{A}, \mathbf{E}) = \int_0^1 (t(\mathbf{A} - \mathbb{I}) + \mathbb{I})^{-1} \mathbf{E} (t(\mathbf{A} - \mathbb{I}) + \mathbb{I})^{-1} dt, \quad (47)$$

$$r(\hat{\mathbf{A}}) = -\hat{\mathbf{A}} - i\omega\mathbb{I}. \quad (48)$$

The function $L(\mathbf{A}, \mathbf{E})$ can be efficiently evaluated using the algorithm by [5].

By substituting (46) into (44) and using (47) with the fact that we can interchange the tr-operator and the integral we obtain

$$\begin{aligned} \left[\frac{\partial \|E\|_{\mathcal{H}_{2,\omega}}^2}{\partial \hat{\mathbf{A}}} \right]_{ij} &= 2 \text{tr} \left[\frac{\partial \hat{\mathbf{A}}^\top}{\partial \hat{a}_{ij}} \left(\mathbf{Y}_\omega^\top \mathbf{X} + \hat{\mathbf{Q}}_\omega \hat{\mathbf{P}} \right) + \frac{\partial \hat{\mathbf{S}}_\omega^*}{\partial \hat{a}_{ij}} \left(\hat{\mathbf{C}}^\top \hat{\mathbf{C}} \hat{\mathbf{P}} - \hat{\mathbf{C}}^\top \mathbf{C} \mathbf{X} \right) \right] \\ &\quad - 2 \text{tr} \left[\frac{\partial \hat{\mathbf{S}}_\omega}{\partial \hat{a}_{ij}} \left[\hat{\mathbf{B}} \left(\mathbf{D}^\top - \hat{\mathbf{D}}^\top \right) \hat{\mathbf{C}} \right] \right] \\ &= 2 \text{tr} \left(\frac{\partial \hat{\mathbf{A}}^\top}{\partial \hat{a}_{ij}} \left(\mathbf{Y}_\omega^\top \mathbf{X} + \hat{\mathbf{Q}}_\omega \hat{\mathbf{P}} \right) \right) - 2 \text{tr} \left(\frac{\partial \hat{\mathbf{A}}^\top}{\partial \hat{a}_{ij}} \text{Re} \left[\frac{i}{\pi} L \left(r(\hat{\mathbf{A}}), \mathbf{V} \right) \right]^\top \right) \\ &= \text{tr} \left(\frac{\partial \hat{\mathbf{A}}^\top}{\partial \hat{a}_{ij}} \left[2 \left(\mathbf{Y}_\omega^\top \mathbf{X} + \hat{\mathbf{Q}}_\omega \hat{\mathbf{P}} \right) - 2\mathbf{W} \right] \right), \quad (49) \end{aligned}$$

where

$$\mathbf{W} = \text{Re} \left[\frac{i}{\pi} L \left(r(\hat{\mathbf{A}}), \mathbf{V} \right) \right]^\top, \quad (50)$$

$$\mathbf{V} = \hat{\mathbf{C}}^\top \hat{\mathbf{C}} \hat{\mathbf{P}} - \hat{\mathbf{C}}^\top \mathbf{C} \mathbf{X} - \hat{\mathbf{C}}^\top \left(\mathbf{D} - \hat{\mathbf{D}} \right) \hat{\mathbf{B}}^\top. \quad (51)$$